# Data Assimilation Schemes in Colombian Geodynamics - Cooperative Research Plan for 2017 - 2020 Between Universidad EAFIT and TUDelft, With the Help of Universidad de Antioquia and universidadNacional de Colombia Sede Medellin

Start date: 1 January 2017
End date: 30 December 2020

## Technical Report – Data Assimilation On Air Quality Models For Aburrá Valley

**Medellin Air qUalityInitiative MAUI**

MAUI-RT-03

| | |
|---|---|
| Executingagency | Universidad EAFIT<br>Cra 49 No 7sur - 50<br>Medellín, Colombia |
| | Grupo de investigación en modelado matemático – GRIMMAT<br>Grupo reconocido por COLCIENCIAS Categoría A<br>Grupo de investigación en Biodiversidad, Evolución y Conservación - BEC |
| Responsible | Prof. Olga Lucia Quintero Montoya<br>Prof. Nicolás Pinel Peláez.<br>Investigadores |
| Cooperatingentities | Department of Applied Mathematics - Tu Delft, Delft The Netherlands<br>TNO |
| Responsible | ArnoldHeemink<br>ArjoSegers |

## EDITION AND DISTRIBUTION CONTROL

| Edition | Control* | Nombre y Cargo | Signature | Entity | Date (DD/MM/YYYY) |
|---|---|---|---|---|---|
| 1 | Creation | Santiago López Restrepo | | Universidad EAFIT | 17/04/2017 |

* (Creation - Revision - Modification - Distribution)

**CONTENTS**

**INTRODUCTION** ............ ¡Error! Marcador no definido.
**THEORETICAL FRAMEWORK** ............ ¡Error! Marcador no definido.
AIR POLLUTANTS ............ ¡Error! Marcador no definido.
LOTOS-EUROS MODEL ............ ¡Error! Marcador no definido.
DATA ASSIMILATION ............ ¡Error! Marcador no definido.
VARIATIONAL METHODS ............ ¡Error! Marcador no definido.
ENSEMBLE KALMAN FILTERING ............ ¡Error! Marcador no definido.
**RESEARCH QUESTION** ............ ¡Error! Marcador no definido.
**OBJECTIVES** ............ ¡Error! Marcador no definido.
EXPECTED RESULTS ............ ¡Error! Marcador no definido.
**REFERENCES** ............ ¡Error! Marcador no definido.

**FIGURES LIST**

Figure 1. Variational Method 4D-Var (Lahoz & Schneider, 2014). .......................... **¡Error! Marcador no definido.**
Figure 2. MPC block diagram (Shi, Kelkar, & Soloway, 2005). ............................... **¡Error! Marcador no definido.**
Figure 3. MPC block diagram for the problem of air quality in the Aburrá Valley. ... **¡Error! Marcador no definido.**

## 1. INTRODUCTION

Air pollution is defined as the presence of solid, liquid or gaseous components in the atmosphere. Above all, in the troposphere (lower atmosphere layer in contact with the land) the pollutants are a risk and trouble for living beings or goods in general. In fact, air pollution is the major environmental problem in modern human history (Green & Sánchez, 2012). Nowadays environmental pollution can be produced by natural or human actions; as evidence of natural sources are mainly forest fires, volcanic emissions, dust, sand, vegetation and wildlife; and main human sources of air pollution for instance are industry, power generation, transportation, deforestation and cattle raising (Borrego et al., 2015).

The city of Medellín is located in the center of Aburra Valley, and has been one of the most polluted cities in Latin America, where its geographic qualities, the plenty industry and the growing car fleet provide the poor air quality (Green & Sánchez, 2012). A proof of this is given around March of 2016 when the environmental pollution ratings were the highest registered in the all history of the city and several factors that influenced these ratings were the ENSO (commonly called 'El Niño'), little rains, weak winds and increased temperatures, generating accumulation of pollutants from fuel combustion, Sahara sand and smoke from forest fires in Colombia and Venezuela, that joined with kept off the scattering of pollutants and raised up the concentration levels of pollutants like PM2.5 and PM10 (Alsema, 2016). Therefore the measures registered were PM2.5 higher than 160 µg/m3 24-hour, when the guideline of World Health Organization (WHO) is 25 µg/m3 24-hour mean (Ospina, 2016).

Due to the magnitude of the problem that air pollution has become, many efforts have been made to monitor, reduce and prevent the spread of pollutants in the air. As a first containment action is greatly important to know the pollution concentrations and the air quality in an area and the time given. For this, nowadays, an advanced system of measure and mathematical models exist, which represent the air pollution dynamics. These mathematical models known as Air Quality Models (AQM), allow a permanent monitoring and in many cases predictions of the air quality behavior.

In view of the problem that air pollution represents to the city of Medellín, is necessary looking for mechanisms that allow contain and reduce levels of pollutants in the environment. The first step to decrease these pollutants is to know their behavior and the air state of the city. Having measures of the main air pollutants and knowing their behaviors in the Aburra Valley atmosphere could be proposed actions that help to improve air quality levels and reduce impacts of pollution on the population. Without these factors in advance is impossible to think of responses or preventives actions. Following the preceding outlook has been proposed the next question:

Is it possible to model the atmosphere behavior at the Aburra Valley scale so that it can to monitor environmental pollutants and predict their behavior?

## 2. THEORETICAL FRAMEWORK

### 2.1. AIR POLLUTANTS

In our days, air pollution is listed as the highest environmental risk for human health. In accordance with a report of the World Health Organization (WHO), in 2012 one of each nine deaths were consequence of conditions related to air pollution (World Health Organization, 2016). Besides the principal health risk associated with air pollution are respiratory and skin (dermatological) diseases. Respiratory disease is one of the main causes of natural deaths in European, American and Asian countries (World Health Organization, 2016). Additionally, different air pollutants like Carbon Dioxide ($CO_2$), Ozone ($O_3$) and coal ash contribute substantially to climate change, another current wide environmental risk (Sauter et al., 2012). In the same way in Latin America and the Caribbean at least 100 million people are exposed to air pollution above World Health Organization recommended levels, which means they are in high risk of public health. Meanwhile, the biggest cities in Latin America keep alarming pollution levels, because they gather the majority of industries and combustion vehicles. The 10 most polluted cities in Latin America are: Monterrey, Guadalajara and the Federal District (México), Cochabamba (Bolivia), Santiago (Chile), Lima (Peru), Bogotá and Medellín (Colombia), Montevideo (Uruguay) and San Salvador (El Salvador). The leading air pollutants are shown below (Green & Sánchez, 2012):

**Particulate matter PM2.5 y PM10:** is made up of a mix among tiny solid and liquid particles with lesser sizes than 2.5 and 10 micrometers, respectively. This kind of particles is easily breathed in humans, bringing about serious respiratory diseases and these being composed of chemical substances can result carcinogenic.

**Ozone ($O_3$):** is a gas made up of three oxygen molecules. Although be essential for life, which protect the surface from ultraviolet rays emitted by the sun (designated ozone layer which is located in the stratosphere) has adverse effects in the respiratory system even at relatively low levels. Ozone is not directly released to the troposphere, it is formed by photochemical reactions such as the Oxides of Nitrogen ($NO_x$) and other volatile chemical components.

**Nitrogen Dioxide ($NO_2$):** is a gas formed as a result of reaction between Oxygen and Nitrogen. It is mainly produced of fossil fuels in transportation, industry such as power generation. The $NO_2$ at higher concentrations cause irritation in airways of the lungs, increasing the risk and respiratory diseases. It also contributes to the formation of ozone in the troposphere.

**Sulfur dioxide ($SO_2$):** similar to the $NO_2$ it is a gas formed during combustion of fossil fuels like oil and coal. The exposition at higher concentrations cause respiratory and heart diseases.

### 2.2. LOTOS-EUROS MODEL

In the last 20 years, Air Quality Models have seen a great growth and development; in consequence diversity of models exist, differing in their complexity, size of the region in study, to the method used for its development (Thunis et al., 2016). Very general, the Air Quality Models can be broken up in four categories according with their dynamical behavior: i) Gaussian, ii) statistic, iii) Lagrangian and iv) Eulerian like chemistry transport models (Lateb et al., 2016; Thunis et al., 2016); the latter being the most used and reported for monitoring and predicting the pollution behavior and define the air quality in bigger areas. So, these are frequently used in areas with sizes like countries or continents and have been less used in areas like cities.

One of the most used and studied Air Quality Models in the present is the LOTOS-EUROS (Mues et al., 2014). The LOTOS-EUROS (LOng Term Ozone Simulation- EURopean Operational Smog model) is a chemical transport model that models in three dimensions the air pollution in the lower troposphere. This model was developed in 2004 by TNO and RIVM/MNP organizations, in Netherlands, unifying the previous developed LOTOS and EUROS models. At the beginning it was developed like a model focused on ozone, but actually, the LOTOS-EUROS (versión 1.8) allows calculate concentrations of Ozone, Particulate Matter, Nitrogen Dioxide, heavy metals and organic pollutants with a standard model resolution of approximately 36x28 km. (Sauter et al., 2012).

The LOTOS-EUROS has widely used in different projects located around the world, whereby it shows the capacity of the model. As well, it is within the framework of the project MACC II, that is looking to produce the forecast at European continent level in air quality, meteorology and solar radiation (Marécal et al., 2015). The MACC II project uses the network of satellites and sensors denominated COPERNICUS along with LOTOS-EUROS (among other air quality models) to make the predictions of air quality. Likewise, the LOTOS-EUROS is used in Netherlands to predict Ozone concentrations and PM in national territory. This project is named SmogProg and is used by Dutch authorities as official forecasts, it is directly published to the institutions and population in the country (Hendriks et al., 2013). The LOTOS-EUROS model has not been only implemented in Europe, actually is part of the project PANDA, that collect a set of models and looks for modeling and predicting pollutants concentrations in Chinese territory. Similarly, in the north of Africa is settled the Regional Center for Northern Africa, Middle East and Europe, which uses the LOTOS-EUROS to monitoring and predicting the air quality in Northern Africa, Middle East and Eastern Europe. In America continent the model has been implemented by Brazil to monitoring and predicting Ozone concentrations, Nitrogen Dioxide and PM 2.5 while was taking place the 2014 FIFA World Cup.

The dynamics of the pollutants in the model LOTOS-EUROS is regulated by processes of chemical reactions, diffusion, drag, dry deposition and wet, emissions and aversion (Sauter et al., 2012; van Loon, Builtjes, & Segers, 2000). The LOTOS-EUROS dynamic is given by:

$$\frac{\partial C}{\partial t}+U\frac{\partial C}{\partial x}+V\frac{\partial C}{\partial y}+W\frac{\partial C}{\partial z}=\frac{\partial}{\partial t}\left(K_h\frac{\partial C}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_h\frac{\partial C}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_z\frac{\partial C}{\partial z}\right)+E+R+Q-D-W \tag{1}$$

when $C$ the concentration of a pollutant, $U$, $V$ and $W$, are wind components in West-East, South-North direction and vertical direction, respectively, $K_h$ and $K_z$ are horizontal and vertical coefficients by diffusion of turbulence. $E$ represents the entrainment or detrainment due to variations in layer height. $R$ represents generation and consumption rates of pollutant by chemical reactions. $Q$ is contribution by emissions, $D$ and $W$ are loss by dry and wet deposition process, respectively.

The main equation of LOTOS-EUROS dynamic is composed by different operators, each one models different components of pollutants behavior. The operators that compose the LOTOS-EUROS are: i) the transport operator, ii) the chemistry operator, iii) the dry deposition operator and iv) the wet deposition operator. Emissions and values related with meteorology are directly taken from data sources as satellite or measuring devices located in the land surface.

The transportation operator consists in the dynamic of advection in three dimensions, horizontal and vertical diffusion and entrainment. The horizontal advection is described by horizontal winds (*U,V*) that are inputs to the model. The vertical wind component (*W*) is calculated by the model through the convergence and divergence of the horizontal winds. The horizontal diffusion coefficient ($K_h$) is calculated through an empiric constant η and the speed tensor deformation *Def*, as shown in (2) and (3):

$$K_h = \eta |Def| \tag{2}$$

$$|Def| = \sqrt{\left[\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial x} - \frac{\partial V}{\partial y}\right)^2\right]} \tag{3}$$

The chemistry operator models everything related to the production and consumption of components by different chemical reactions in the atmosphere. Due to the complexity of LOTOS-EUROS, to handle a complete mechanism of chemical reactions could cause an unmanageable model. To avoid this problem, the LOTOS-EUROS can use one of two mechanisms of simplified reactions, Carbon Bond-IV (CB-IV) or CB99. The CB-IV uses nine primary components directly issued to the atmosphere and a total of 81 reactions to determine secondary species produced in the atmosphere. The second mechanism belonging to the LOTOS-EUROS is the CB99, which is a variation of the CB-IV. The CB99 uses 42 chemical species and 95 reactions, including 13 photolytic reactions.

The dry deposition is divided in two phases, the dry deposition of gases and the dry deposition of particles. The dry deposition of gases is modeled through the transfer of gases between the land surface and the atmosphere, result of the difference in concentrations and resistance between them. In the dry deposition of particles the scheme used depend of the given use to the land over is made the analysis. This scheme confer flexibility and dynamism in the aerosol size, although due to simplicity is taken two sizes of reference 0.7 and 8.0 µm.

The operator of wet deposition is modeled through the belowcloud scavenging process. The belowcloud scavenging process uses a sweep coefficient Λ [$s^{-1}$] that describes the mass transfer speed of a pollutant from the air to the raindrops. The value of the sweep coefficient depends of the considered component. Nevertheless, in general the decrease in concentration C [µg / m3] is calculated as:

$$\frac{\partial C}{\partial t} = -\Lambda C \tag{4}$$

The contribution to the flow of wet deposition Δ*D* [µg/$m^2$] is described by:

$$\Delta D = C_0(1 - e^{-\Lambda t})\Delta z$$

where $C_0$ is the initial concentration and $\Delta z$ [m] is the height of a cell in the resolution grid.

The LOTOS-EUROS model is considered on a large scale due to the solution of (1) is executed for different components and in each point belonging to a grid on the region of analysis. Because of this process the vector of states has large size (in the order of thousand).

## 2.3. DATA ASSIMILATION

Data Assimilation is a mathematical process that provides integration between measured values (observations) and a dynamical transport model, to improve the operation of the model. The output value provided by the model has a smaller error than the output value provided by the model without observations. Data Assimilation has two key objectives, to improve the operation in predictions of model states and estimate unknown parameters of the model (Berardi, Andrisani, Lopez, & Vurro, 2016). The Data Assimilation has been proven in different science fields as oceanography, climatology, air quality models and atmospheric chemistry (van Loon et al., 2000). With the Data Assimilation are filled voids of time and space where it have not observations, alike, there is added a value to the model, restricting it with the observations, and in the same way is possible to make a consistent simulation and lead the surface area or study region. Data Assimilation allows integrate models and observations with different scales of size and temporal sampling (Lahoz & Schneider, 2014).

When two sources of information are combined, Data Assimilation assumes that both the model and the measurements are subject to errors. These errors are impossible to know with accuracy and need to be specified in statistical and probabilistic terms. Data assimilation is not only looking to reduce the model error in points or time with observations, its mission is to digest the observation based on the laws given by the model and to determine the dynamic evolution of the model state that represent better the measurements (Bocquet et al., 2015; van Loon et al., 2000).

According to the implemented method in Data Assimilation, exist two main categories: filtering methods and Variational methods (Lahoz & Schneider, 2014; Lu, Lin, Heemink, Fu, & Segers, 2016). The filtering methods, also called the Kalman Filter, are a type of sequential method that looks for improvements in the prediction of the model reducing the covariance between observations and model outputs. For this methods are used variations of Kalman Filter that make it more efficient in large scale problems (Berardi et al., 2016; Sebacher, Hanea, & Heemink, 2013). On other hand, Variational methods looks for the optimal states set that minimize cost functions between observations made and model outputs (Altaf, El Gharamti, Heemink, & Hoteit, 2013; Lu et al., 2016).

### 2.3.1. VARIATIONAL METHODS

One of the most widely used variational methods in Data Assimilation is the method 4D-Var, as its name implies, the method 4D-Var is a variational method in four dimensions. It is an extension of the method 3D-Var generalized with the incorporation of time as a fourth variable. The method 4D-Var looks for initial conditions of the model states which can provide a better fit with respect to observations within the interval of assimilation (Altaf et al., 2013; Lu et al., 2016). Figure 1 shows an illustrative explanation of the method.

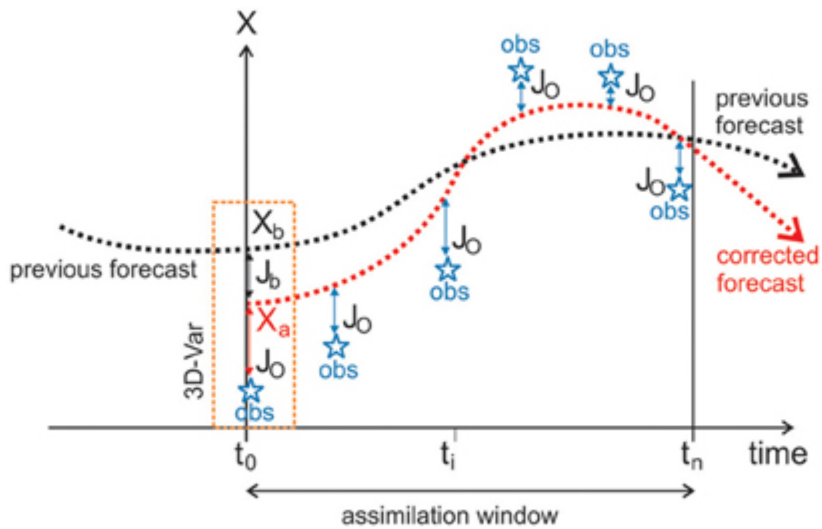


*Figure1. Variational Method 4D-Var* (Lahoz & Schneider, 2014).

The method 4D-Var look for minimize the cost function that relates the difference among the current state, the previous states and the prediction of the model from observations in interval of assimilation. Consider the discrete model of a nonlinear dynamical system given by (5) and (6) (Lu et al., 2016):

$$X(t_{i+1}) = M_i\big(X(t_i), U(t_i)\big) \tag{5}$$

$$Y(t_i) = H\big(X(t_i)\big) + \eta(t_i) \tag{6}$$

where, $X(t_i)$ is a vector of states in the time $i$, $U(t_i)$are inputs of the system, $M_i$ is a nonlinear operator which propagates the states, $Y(t_i)$ are model observations, $H$ is a nonlinear operator which maps the state space on observation space and $\eta(t_i)$ is the Gaussian white noise with cero average and standard deviation given by the matrix $R_i$.

The method 4D-Var minimizes the functional cost (7) using the initial state value as decision variable (Altaf et al., 2013).

$$J(X_0) = \frac{1}{2}\left(X^b - X_0\right)^T B_o^{-1}\left(X^b - X_0\right) + \frac{1}{2}\sum_{i=0}^{N}\Big(Y(t_i) - H\big(X(t_i)\big)\Big)^T R_i^{-1}\Big(Y(t_i) - H\big(X(t_i)\big)\Big) \tag{7}$$

where $X^b$ is the *a priori* state to$X_0$, assuming a correlation with covariance matrix $B_0$and $N$ is the assimilation window size.

**2.3.2. ENSEMBLE KALMAN FILTERING**

The Ensemble Kalman Filter (EnKF) has been developed by(Evensen, 1994). It has been implemented in many studies because of its simple formulation and its easy implementation, because is not necessary derivation and integrations backward on time(Roop, Vyatkin, & Salcic, 2007). The EnKF was designed to resolve two major problems related to the use of the Extended Kalman Filter (EKF) with nonlinear dynamics in large state spaces. The mainly problems are that EKF uses a closure scheme when third- and higher-order moments in the error covariance equations are discarded, and the other problem is the huge computer requirements. The EnKF is a sequential filter method, it means that the model is integrated forward on time, and when a measurement is available, is used to reinitialize the model before the integration (Evensen, 2003). To understand the EnKF first, is showed the normal Kalman Filter. Consider the linear state-space system showed in (8).

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{M}_t \boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \qquad && \boldsymbol{w}_t \sim N_n(0, \boldsymbol{Q}_t) \\ \boldsymbol{y}_t &= \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{v}_t, && \boldsymbol{v}_t \sim N_{m_t}(0, \boldsymbol{R}_t) \end{aligned} \tag{8}$$

where $\boldsymbol{x}_t$ is the state vector, $\boldsymbol{y}_t$ observed data vector, $\boldsymbol{M}_t, \boldsymbol{Q}_t$ are the transition matrix and model error covariance and $\boldsymbol{H}_t, \boldsymbol{R}_t$are the observation matrix and observation error covariance respectively. The filtering problem involves estimation of the state $\boldsymbol{x}_t$ conditional on$\boldsymbol{y}_t$. Consider the filtering distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ is Gaussian, the moments are obtained by the Kalman filter. Assume that the filtering distribution at time $t-1$ is:

$$\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1} \sim N_n(\mu_{t-1}^a, \boldsymbol{P}_{t-1}^a) \tag{9}$$

The first step of the Kalman Filter methods is the forecast step, in this step, the filter calculates the moments of the distribution at time t with the data at time $t-1$.

$$\boldsymbol{x}_t|\boldsymbol{y}_{t-1} \sim N_n(\mu_t^f, \boldsymbol{P}_t^f) \tag{10}$$

where:

$$\mu_t^f = \boldsymbol{M}_t \mu_{t-1}^a, \qquad \boldsymbol{P}_t^f = \boldsymbol{M}_t \boldsymbol{P}_{t-1}^a \boldsymbol{M}_t' + \boldsymbol{Q}_t \tag{11}$$

The next step called update or analysis step, updates the state at time $t$, with the data in time $t, \boldsymbol{y}_t$.

$$\begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{pmatrix} |\boldsymbol{y}_{t-1} \sim N\left( \begin{pmatrix} \mu_t^f \\ \boldsymbol{H}_t \mu_t^f \end{pmatrix} \begin{pmatrix} \boldsymbol{P}_t^f \boldsymbol{P}_t^f \boldsymbol{H}_t' \\ \boldsymbol{H}_t \boldsymbol{P}_t^f \boldsymbol{H}_t \boldsymbol{P}_t^f \boldsymbol{H}_t' + \boldsymbol{R}_t \end{pmatrix} \right) \tag{12}$$

Using the properties of the multivariate normal distribution, it follows that $\boldsymbol{x}_t|\boldsymbol{y}_t \sim N_n(\mu_t^a, \boldsymbol{P}_t^a)$, where:

(13)

$$\mu_t^a = \mu_{t-1}^f + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{H}_t\mu_t^f), \qquad \boldsymbol{P}_t^a = (\boldsymbol{I}_n - \boldsymbol{K}_t\boldsymbol{H}_t')\boldsymbol{P}_t^f$$

The matrix $\boldsymbol{K}_t$ is the Kalman gain matrix, defined as:

$$\boldsymbol{K}_t = \boldsymbol{P}_t^f\boldsymbol{H}_t'(\boldsymbol{H}_t\boldsymbol{P}_t^f\boldsymbol{H}_t' + \boldsymbol{R}_t)^{-1} \tag{14}$$

The EnKF is an approximate version of the Kalman filter, where the state distribution is represented by a sample from the distribution2(Barbu, Segers, Schaap, Heemink, & Builtjes, 2009; Fu et al., 2016). For an ensemble of states from $i = 1 \ldots N$, the forescat step obtains a sample from the forescast distribution by applying the evolution equation to each posterior ensemble member $\boldsymbol{x}_{t-1}^{a(i)}$ at time $t-1$:

$$\boldsymbol{x}_t^{f(i)} = \boldsymbol{M}_t\boldsymbol{x}_{t-1}^{a(i)} + \boldsymbol{w}_t^{(i)}, \qquad \boldsymbol{w}_t^{(i)} \sim N(0, \boldsymbol{Q}_t) \tag{15}$$

The forecast ensemble $\boldsymbol{x}_t^{f(i)}$ must then be updated based o the new data $\boldsymbol{y}_t$ . This are do stochastically using the perturbed observations. Given the forecasts$\boldsymbol{x}_t^{f(i)}$, we sample forecast observations as:

$$\boldsymbol{y}_t^{f(i)} = \boldsymbol{H}_t\boldsymbol{x}_t^{f(i)} + \boldsymbol{v}_t^{(i)}, \qquad \boldsymbol{v}_t^{(i)} \sim N(0, \boldsymbol{R}_t) \tag{16}$$

According with (16), the update for the ensemble $\boldsymbol{x}_t^{a(i)}$, is given by:

$$\boldsymbol{x}_t^{a(i)} = \boldsymbol{x}_t^{f(i)} + \widehat{\boldsymbol{K}}_t\left(\boldsymbol{y}_{t-}\boldsymbol{y}_t^{(i)}\right) \sim N(\mu_t^a, \widehat{\boldsymbol{P}}_t^a) \tag{17}$$

where:

$$\mu_t^a = E(\boldsymbol{x}_t^{a(i)}) \tag{18}$$

$$\widehat{\boldsymbol{P}}_t^a = \widehat{\boldsymbol{P}}_t^f - \widehat{\boldsymbol{K}}_t\boldsymbol{H}_t\widehat{\boldsymbol{P}}_t^f \tag{19}$$

$$\widehat{\boldsymbol{K}}_t = \widehat{\boldsymbol{P}}_t^f\boldsymbol{H}_t'(\boldsymbol{H}_t\widehat{\boldsymbol{P}}_t^f\boldsymbol{H}_t' + \boldsymbol{R}_t)^{-1} \tag{20}$$

$$\widehat{\boldsymbol{P}}^f = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_t^{f(i)} - \overline{\boldsymbol{x}}_t^f)(\boldsymbol{x}_t^{f(i)} - \overline{\boldsymbol{x}}_t^f)' \tag{21}$$

## 3. RESEARCH QUESTION

The air quality control problem can be represented as a Model Predictive Control (MPC) problem. The MPC uses a model of the process to be controlled to predict the best control action and take de best possible decision(Bai, Xiao, Yang, & Zhang, 2009; Qin & Badgwell, 2000).For this type of controller is very important the accuracy of the model, if the model does not represent the process correctly, the controller will generate an incorrect control action on the process(Oldewurtel et al., 2010). The typical block diagram for a MPC is showed in Figure 2.

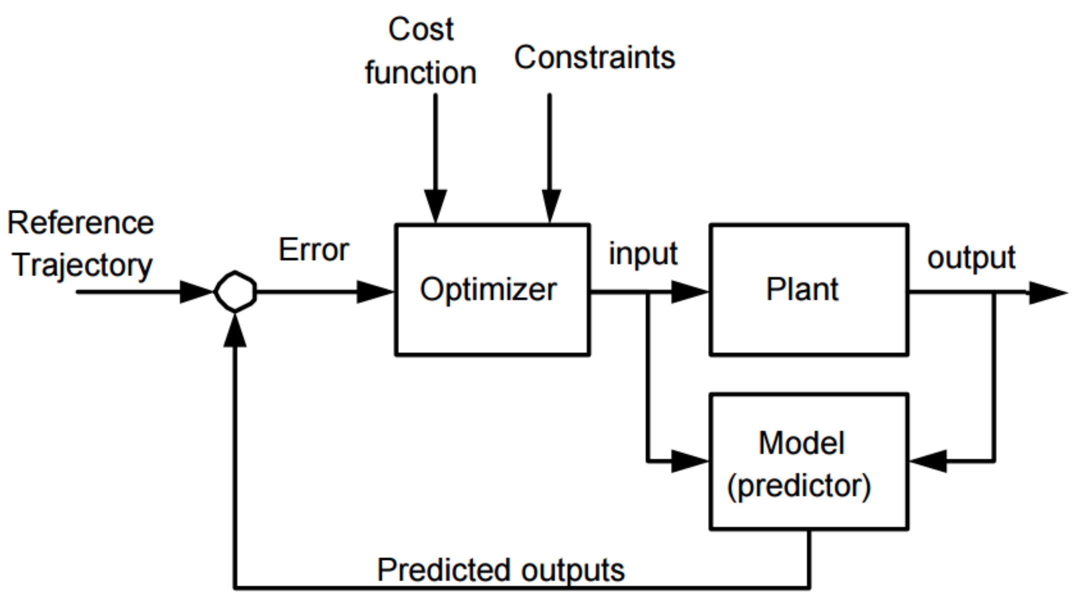


*Figure 2. MPC block diagram* (Shi, Kelkar, & Soloway, 2005)

Extrapolating the concept of MPC to the problem of the air quality in the Aburrá Valley, is presented the diagram block showed in the Figure 3.

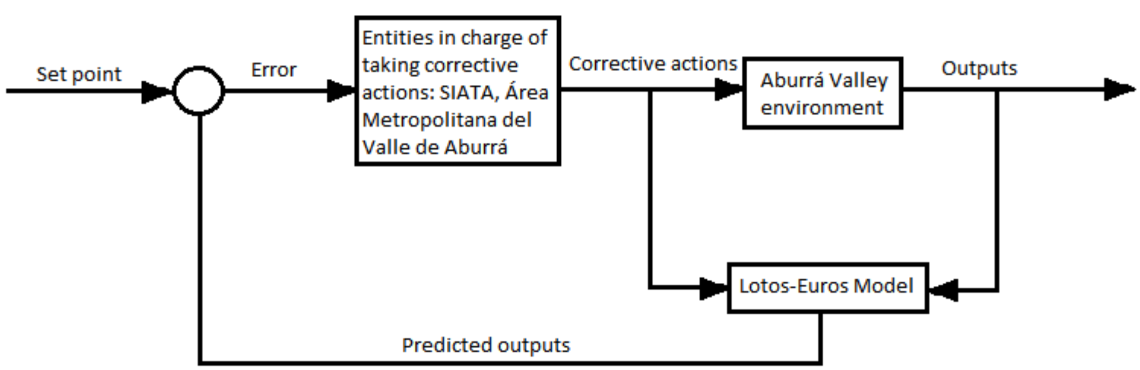


*Figure 3. MPC block diagram for the problem of air quality in the Aburrá Valley.*

For the control problem of air quality, in this case, for the Aburrá Valley, the entities in charge of taking corrective environmental actions (SIATA, Área Metropolitana del Valle de Aburrá and the municipalities) play the role of the controller. These entities would use the LOTOS-EUROS forecast to make the optimum action over the transportation and the industry. This control architecture looks for reduce the differences between the measurements of air quality (outputs) and the local air quality index (Set Point), keeping low the pollutants levels and acting before that represent a public health problem..

Due to the current maximum resolution possible in the model LOTOS-EUROS (9 km x 7 km) there would not be a correct representation of the Aburrá Valley atmosphere, or at least, not enough to execute analysis able to understand the local behavior. Another influential factor is the geography that can produce error in the vertical layers development of model. In other way, observation of concentrations of any component in an air column from satellite data and the assimilation is complex, for this in (Lu et al., 2016) has proposed a method called Trajectory-Based 4D-Var used in cases which the number of parameters is reduced (lesser than 100 parameters to be assimilated).

The Aburrá Valley has SIATA system (due to its acronym in Spanish: Sistema de AlertaTemprana de Medellín y el Valle de Aburrá), which has deployed sensors network in the metropolitan area of Aburrá Valley, these are able to measure different air pollutants. With the previous outlook the following research question has been proposed:

Is it possible obtain a complete dynamic representation of the atmosphere in Aburrá Valley, using LOTOS-EUROS model, a Data Assimilation technique and data provided by satellite and the SIATA system?

**OBJECTIVES**

The main objective of this proposal is: to assimilate the Lotos-Euros in the Aburrá Valley with data measured on surface that allows a complete representation of the pollutants in the low atmosphere. Specific objectives are:

- To determine the viability of the Lotos-Euros model in the Tropical Andes region.
- To reduce the resolution of the Lotos-Euros model to a size that allows to represent the dynamics and behavior of the pollutants in the Aburrá Valley.
- To analyze the different methods of Data Assimilation, from its structure and operation in air quality models under the orographic conditions of the Aburrá Valley.
- To compare the performance of different methods of Data Assimilation for the Lotos-Euros model on the Aburrá Valley.

- To contribute to the field of data assimilation through the development of new schemes from the solution of the particular problem related to air quality in the Aburrá Valley.

- To increase the possibility of forecasting of the Air Quality in Aburrá Valley using Lotos-Euros Model through Data Assimilation taking into account the need for more accurated data provided by governmental agencies such as SIATA and ÁreaMetropolitana.

**EXPECTED RESULTS**

- Formation as PhD in Mathematical Engineering.
- At least one scientific papers.
- Four (4) Reports Working papers.
- An Aburrá Valley Model for Air Quality completely assimilated.

**REFERENCES**


Alsema, A. (2016, March 16). Medellín declares red alert over record air pollution levels. *COLOMBIA REPORTS*. Retrieved from http://colombiareports.com/medellin-red-alert-excessive-air-pollution

Altaf, M. U., El Gharamti, M., Heemink, W., & Hoteit, I. (2013). A reduced adjoint approach to variational data assimilation. *Computer Methods in Applied Mechanics and Engineering*, *254*, 1–13. http://doi.org/10.1016/j.cma.2012.10.003

Bai, J., Xiao, H., Yang, X., & Zhang, G. (2009). Study on integration technologies of building automation systems based on web services. *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, 262–266. http://doi.org/10.1109/CCCM.2009.5267730

Barbu, A. L., Segers, A. J., Schaap, M., Heemink, A. W., & Builtjes, P. J. H. (2009). A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe. *Atmospheric Environment*, *43*(9), 1622–1631. http://doi.org/10.1016/j.atmosenv.2008.12.005

Berardi, M., Andrisani, A., Lopez, L., & Vurro, M. (2016). A new data assimilation technique based on ensemble Kalman filter and Brownian bridges: An application to Richards' equation. *Computer Physics Communications*, *208*, 43–53. http://doi.org/http://dx.doi.org/10.1016/j.cpc.2016.07.025

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Aabkar, R., Carmichael, G. R., … Seigneur, C. (2015). Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. *Atmospheric Chemistry and Physics*, *15*(10), 5325–5358. http://doi.org/10.5194/acp-15-5325-2015

Borrego, C., Coutinho, M., Costa,  a. M., Ginja, J., Ribeiro, C., Monteiro, … Miranda, I. (2015). Challenges for a New Air Quality Directive: The role of monitoring and modelling techniques. *Urban Climate*, *14*, 328–341. http://doi.org/10.1016/j.uclim.2014.06.007.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *99*.

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. http://doi.org/10.1007/s10236-003-0036-9

Fu, G., Heemink, A., Lu, S., Segers, A., Weber, K., & Lin, H. X. (2016). Model-based aviation advice on distal volcanic ash clouds by assimilating aircraft in situ measurements. *Atmospheric Chemistry and Physics*, *16*(14), 9189–9200. http://doi.org/10.5194/acp-16-9189-2016

Green, J., & Sánchez, S. (2012). *Air Quality in Latin America: An Overview*. *Clean Air Institute*. Washington D.C., USA.

Hendriks, C., Kranenburg, R., Kuenen, J., van Gijlswijk, R., Wichink Kruit, R., Segers, A., … Schaap, M. (2013). The origin of ambient particulate matter concentrations in the Netherlands. *Atmospheric Environment*, *69*, 289–303. http://doi.org/10.1016/j.atmosenv.2012.12.017

Lahoz, W. a., & Schneider, P. (2014). Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science*, *2*(May), 1–28. http://doi.org/10.3389/fenvs.2014.00016

Lateb, M., Meroney, R. N., Yataghene, M., Fellouah, H., Saleh, F., & Boufadel, M. C. (2016). On the use of numerical modelling for near-field pollutant dispersion in urban environments − A review. *Environmental Pollution*, *208*, 271–283. http://doi.org/10.1016/j.envpol.2015.07.039

Lu, S., Lin, H. X., Heemink, A. W., Fu, G., & Segers, A. J. (2016). Estimation of Volcanic Ash Emissions Using Trajectory-Based 4D-Var Data Assimilation. *Monthly Weather Review*, *144*(2), 575–589. http://doi.org/10.1175/MWR-D-15-0194.1

Marécal, V., Peuch, V. H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., … Ung, A. (2015). A regional air quality forecasting system over Europe: The MACC-II daily ensemble production. *Geoscientific Model Development*, *8*(9), 2777–2813. http://doi.org/10.5194/gmd-8-2777-2015

Mues, A., Kuenen, J., Hendriks, C., Manders,  a., Segers,  a., Scholz, Y., … Schaap, M. (2014). Sensitivity of air pollution simulations with LOTOS-EUROS to the temporal distribution of anthropogenic emissions. *Atmospheric Chemistry and Physics*, *14*(2), 939–955. http://doi.org/10.5194/acp-14-939-2014

Oldewurtel, F., Parisio, A., Jones, C. N., Morari, M., Gyalistras, D., Gwerder, M., … Wirth, K. (2010). Energy efficient building climate control using Stochastic Model Predictive Control and weather predictions. *American Control Conference*, 5100–5105. http://doi.org/10.1109/ACC.2010.5530680

Ospina, G. (2016, March 27). Medellín está respirando humo. *EL COLOMBIANO*. Retrieved from http://www.elcolombiano.com/antioquia/por-contaminacion-medellin-pierde-su-horizonte-YF3799634

Qin, S. J., & Badgwell, T. A. (2000). An Overview of Nonlinear Model Predictive Control Applications. In C. I. Byrnes (Ed.), *Nonlinear Model Predictive Control* (26th ed., pp. 369–392). Basel: Birkhäuser Basel.

Roop, P., Vyatkin, V., & Salcic, Z. (2007). Synchronous Execution of IEC 61499 Function Blocks Using Esterel. *2007 5th IEEE International Conference on Industrial Informatics*, 1189–1194. http://doi.org/10.1109/INDIN.2007.4384944

Sauter, F., Swaluw, E. Van der, Manders-groot, A., Kruit, R. W., Segers, A., & Eskes, H. (2012). *TNO report TNO-060-UT-2012-01451*. Utrecht, Netherlands.

Sebacher, B., Hanea, R., & Heemink, A. (2013). A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). *Computational Geosciences*, *17*(5), 813–832. http://doi.org/10.1007/s10596-013-9357-z

Shi, J., Kelkar, A. G., & Soloway, D. (2005). Stable Reconfigurable Generalized Predictive Control With Application to Flight Control. *Journal of Dynamic Systems, Measurement, and Control*, *128*(2), 371–378. http://doi.org/10.1115/1.2194076

Thunis, P., Miranda, A., Baldasano, J. M., Blond, N., Douros, J., Graff, A., … White, L. (2016). Overview of current regional and local scale air quality modelling practices: Assessment and planning tools in the EU. *Environmental Science & Policy*, *65*, 13–21. http://doi.org/10.1016/j.envsci.2016.03.013

van Loon, M., Builtjes, P. J. H., & Segers,  a. J. (2000). Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. *Environmental Modelling and Software*, *15*(6–7 SPEC. ISS), 603–609. http://doi.org/10.1016/S1364-8152(00)00048-7

World Health Organization. (2016). *Ambiental air pollution: A global assessment of exposure and burden of* *disease*. Geneva, Switzerland.